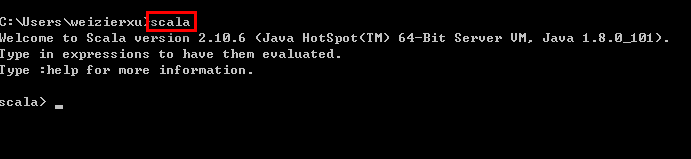
# Spark安装教程

## 1.JDK的安装

## 2. Scala的安装

解压Scala压缩包到指定目录，将Scala安装目录下的bin目录路径，添加到系统变量PATH中。

为了验证是否安装成功，开启一个新的cmd窗口，输入scala然后回车，如果能够正常进入到Scala的交互命令环境则表明安装成功。如下图所示：  


如果不能显示版本信息，并且未能进入Scala的交互命令行，通常有两种可能性：

* Path系统变量中未能正确添加Scala安装目录下的bin文件夹路径名，按照JDK安装中介绍的方法添加即可。
* Scala未能够正确安装，重复上面的步骤即可。

## 3. Spark安装

选择好对应Hadoop版本的Spark版本，下载Spark。

下载完成后将文件进行解压，最好解压到一个盘的根目录下，并重命名为Spark，简单不易出错。并且需要注意的是，在Spark的文件目录路径名中，不要出现空格，类似于“Program Files”这样的文件夹名是不被允许的。

解压后基本上就差不多可以到cmd命令行下运行了。但这个时候每次运行spark-shell（spark的命令行交互窗口）的时候，都需要先cd到Spark的安装目录下，比较麻烦，因此可以将Spark的bin目录添加到系统变量**PATH**中。例如我这里的Spark的bin目录路径为D:\Spark\bin，那么就把这个路径名添加到系统变量的**PATH**中即可，方法和JDK安装过程中的环境变量设置一致，设置完系统变量后，在任意目录下的cmd命令行中，直接执行spark-shell命令，即可开启Spark的交互式命令行模式。

## 4. Hadoop安装

系统变量设置后，就可以在任意当前目录下的cmd中运行spark-shell，但这个时候很有可能会碰到各种错误，这里主要是因为Spark是基于Hadoop的，所以这里也有必要配置一个Hadoop的运行环境。

下载并解压到指定目录，然后到环境变量部分设置**HADOOP\_HOME**为Hadoop的解压目录，我这里是D:\hadoop，然后再设置该目录下的bin目录到系统变量的**PATH**下，我这里也就是D:\hadoop\bin，如果已经添加了**HADOOP\_HOME**系统变量，也可以用%HADOOP\_HOME%\bin来指定bin文件夹路径名。这两个系统变量设置好后，开启一个新的cmd，然后直接输入spark-shell命令。

正常情况下是可以运行成功并进入到Spark的命令行环境下的，但是对于有些用户可能会遇到空指针的错误。这个时候，主要是因为Hadoop的bin目录下没有winutils.exe文件的原因造成的。

下载winutils.exe后，将这个文件放入到Hadoop的bin目录下，我这里是D:\hadoop\bin。

在打开的cmd中输入：**（管理员身份运行）**

D:\hadoop\bin\winutils.exe chmod 777 /tmp/hive



关闭命令行窗口，运行新的命令行窗口：

